



STATISTICAL PRIMER

State Center For Health Statistics

Division of Health Services
Department of Human Resources
P.O. Box 2091, Raleigh, N.C. 27602

Charles J. Rothwell, SCHS Director
George C. Robertson, **PRIMER** Editor

PROBLEMS WITH POPULATION BASES

The intent of the Statistical Primer is to refresh the memory of public health professionals concerning statistical methods. The 1980 census has recently been completed and was the largest peacetime government effort in U.S. history. Since census results affect the reporting of most social statistics, we feel that an examination of census data warrants a place in this series.

Population data, together with percentages and rates based upon them, are invaluable to a wide range of business, marketing and professional interests. In the U.S., these data are largely obtained or estimated from citizens' responses to the U.S. Census which has been taken regularly every 10 years since 1790.

Beginning as a simple list of heads of households with a count of members in five mutually exclusive categories, the census today represents an inventory of many of the personal, social and economic characteristics of the American people. Such an inventory allows us to count and to compute rates for many subgroups of the population. As later described, census counts and certain other enumerations also result in intercensal estimates for a number of population subgroups. These data in turn are used directly or indirectly in many of the estimating procedures to be described in this Primer series; for example, they are the denominators for the crude and adjusted rates described in the preceding Primer. It is thus appropriate that users of this series be aware of weaknesses associated with population data.

Beginning in 1932, the Division of Health Services has annually produced population-based rates like birth and death rates or marriage, divorce and morbidity rates. As before, these rates are valuable to a wide range of interests; however, they are also subject to substantial error. This is due in large measure to the sometimes tenuous population bases that go into them.

Of course, population counts from the U.S. Census are beyond reproach, right? Wrong! For as we all know by now, the Bureau of the Census (BOC) has been bombarded with complaints of undercounts in 1980 and is now in court defending itself against numerous lawsuits. Whether or not these complaints concerning 1980 counts prove to be founded, it is now known that undercounts did occur in 1970.

The Census in North Carolina

Informally, we understand that, even after several post-censal revisions, undercounts in 1970 involved relatively many North Carolinians, especially nonwhite citizens. Hence, since intercensal population estimation and projection procedures use census counts, statewide population bases for the decade of the seventies were low, meaning that corresponding population-based rates were artificially high, especially for nonwhites. Unfortunately, this

finding cannot be generalized to age, sex and many other population subgroups at this time, which serves to leave the trend analyst somewhere in limbo or thereabouts.

But, according to our sources, the good news is that North Carolina's 1980 counts look relatively good. Although Zebulon and Carrboro have filed suits against the BOC, state and county figures for 1980 generally are felt to be much closer to the truth than in 1970. Hopefully, then, adjustments to the 1970 counts will be forthcoming in order that more accurate bases for intercensal years can be generated and the trend analyst can resume his work.

Unfortunately, there is the further bad news that all of this takes time, and in the meantime, we have some questionable population bases for the 1970s. Obviously, then, 5-year rates--so crucial to the study of small populations--should not be computed since they presently involve the latter years of the questionable seventies. Thus, less stable single-year rates must suffice for a time while the trend analyst remains hard-pressed to decide just what is going on.

In addition to the problem of 1970 undercounts, the reader should also be aware of the BOC's revised treatment of Hispanics whereby Mexicans, Puerto Ricans and other Hispanics reporting race in the "other" category were counted as "white" in 1970 but "nonwhite" in 1980. For the entire state, it is estimated that this change involved some 19,000 persons in 1980 who represent only 0.4% or 1.3% respectively of the state's white and nonwhite counts. The situation may be more serious in certain counties, however.

Population Projection/Estimation

Even without errors in census data, population estimation/projection for inter- and post-censal years is a precarious undertaking. In North Carolina, this process involves annual county-specific enumerations of school enrollment in grades 1-8, births and deaths by race, auto and truck registrations, Medicare enrollment, the population in institutions having 200 or more group quarters, and personnel of major military bases. Even if none of these enumerations were subject to error, which they are, the computing formula may be less refined than required, as indeed, North Carolina's formula was found to be in 1975. This, coupled with the 1970 undercounts, suggests that all counties may need to recompute rates for the 70s after final adjustments to the 1970 census allow for new population bases. Similarly, where population bases for the 1960s resulted from interpolation between the 1960 and 1970 censuses, new population bases and rates for the 60s may need to be obtained.

The tables on pages 4 and 5 allow the user to compare total, white and nonwhite county projections (Table 1) to corresponding census counts (Table 2) for April 1, 1980. Note that total counts are final for purposes of BOC publications; white and nonwhite counts are preliminary.

Total projections are actually above census counts for 25 counties while being 10% or more below census counts for only 4 counties--Clay, Dare, Montgomery and Moore. On this basis, the 1980 county projections would not appear grossly biased by undercounts in 1970, but again, this finding cannot be generalized to other population subgroups or to the early years of the 70s. To illustrate, comparisons of the race-specific data of Tables 1 and 2 reveal reasonably close agreement between 1980 projections and 1980 census counts for

whites. Projections were 10% or more below census counts in only 6 counties: Caswell, Clay, Dare, Montgomery, Moore and Perquimans.

But the situation for nonwhites appears far more serious with 33 counties having 1980 projections 10% or more below preliminary census counts. The result is artificially high rates as demonstrated in Table 3 (page 6) where the 33 counties' nonwhite death rates based on projections are compared to those based on census counts. Note that these rates are provisional in numerator counts as well as denominator which are for April 1 rather than the usual July 1.

Other counties should similarly compute both their projection-based and census-based 1980 nonwhite rates to assess changes due to a low or high projection, and all counties should likewise assess changes in their white and total rates. Whatever event is used (deaths, births, etc.), the percentage change is applicable to all other rates based on the same denominator.

The Publication Crunch

For the health data analyst in North Carolina, the upshot of the foregoing litany of problems is a revised agenda for two annual SCHS publications. The revised agenda results from two major problems: (1) unsatisfactory pre-1980 population bases as discussed here and (2) revisions in the International Classification of Diseases which preclude the combining of pre-1979 cause-specific mortality data with those for later years. Decisions relative to this agenda were made by the Division of Health Services' Advisory Committee on Statistical Analysis following deliberations regarding both the accuracy and the timeliness of vital statistics rates. Although none of the chosen alternatives is entirely satisfactory in these respects, the following agenda was felt to represent the best plan at this time:

- The 1979 N.C. Vital Statistics, Volume 2 ("Leading Causes of Mortality") will show single-year numbers and unadjusted rates using population projections. No 5-year rates, adjusted rates or maps will be included. Changes in the cause-of-death classification will be discussed in depth.
- The 1980 N.C. Vital Statistics, Volumes 1 and 2, will show single-year numbers and population-based rates with the 5-year birth-based rates of Volume 1 to remain intact. These volumes will compare rates based on 1980 projections to rates based on 1980 census data (both population bases updated to July 1, 1980). The impact of errors in pre-1980 projections and estimates will be discussed.
- 1979 and 1980 "mentioned conditions" data from death certificates will be treated separately.
- The 1981 Volume 1 will resume standard content if corrected pre-1980 population bases are available. Otherwise, 3-year (1979-81) rates will replace 5-year

Table 1
PROJECTED POPULATION BY RACE: NORTH CAROLINA AND EACH COUNTY, APRIL 1, 1980

<u>Area</u>	<u>Total</u>	<u>White</u>	<u>Nonwhite</u>	<u>Area</u>	<u>Total</u>	<u>White</u>	<u>Nonwhite</u>
North Carolina	5,713,416	4,380,621	1,332,795				
Alamance	100,010	82,007	18,003	Johnston	67,718	54,289	13,429
Alexander	23,439	21,407	2,032	Jones	9,555	5,801	3,754
Alleghany	9,158	8,952	206	Lee	35,739	27,576	8,163
Anson	24,241	13,339	10,902	Lenoir	60,899	36,693	24,206
Ashe	20,682	20,482	200	Lincoln	39,865	36,093	3,772
Avery	14,659	14,564	95	McDowell	35,015	33,445	1,570
Beaufort	40,641	26,577	14,064	Macon	19,834	19,549	285
Bertie	21,296	9,766	11,530	Madison	17,476	17,368	108
Bladen	29,922	18,520	11,402	Martin	25,599	14,784	10,815
Brunswick	36,381	27,639	8,742	Mecklenburg	395,614	291,466	104,148
Buncombe	154,880	141,014	13,866	Mitchell	14,190	14,152	38
Burke	66,116	61,443	4,673	Montgomery	19,585	14,705	4,880
Cabarrus	80,791	68,527	12,264	Moore	45,356	35,397	9,959
Caldwell	62,209	58,277	3,932	Nash	69,159	46,474	22,685
Camden	5,855	3,786	2,069	New Hanover	102,807	77,774	25,033
Carteret	38,918	34,573	4,345	Northampton	23,182	8,964	14,218
Caswell	19,741	10,457	9,284	Onslow	120,658	97,514	23,144
Catawba	103,884	94,582	9,302	Orange	73,833	60,531	13,302
Chatham	30,958	22,137	8,821	Pamlico	9,804	6,585	3,219
Cherokee	18,081	17,501	580	Pasquotank	29,161	18,218	10,943
Chowan	11,951	6,831	5,120	Pender	22,580	14,386	8,194
Clay	5,770	5,703	67	Perquimans	8,926	5,222	3,704
Cleveland	80,488	64,811	15,677	Person	27,402	19,639	7,763
Columbus	53,027	35,879	17,148	Pitt	82,078	54,605	27,473
Craven	71,217	51,282	19,935	Polk	13,117	11,789	1,328
Cumberland	246,628	171,552	75,076	Randolph	86,886	81,383	5,503
Currituck	11,432	9,461	1,971	Richmond	42,334	30,217	12,117
Dare	11,126	10,340	786	Robeson	97,567	39,263	58,304
Davidson	104,553	95,049	9,504	Rockingham	78,362	63,372	14,990
Davie	23,067	21,009	2,058	Rowan	93,783	79,023	14,760
Duplin	40,795	26,929	13,866	Rutherford	52,548	47,391	5,157
Durham	148,164	92,162	56,002	Sampson	49,945	31,672	18,273
Edgecombe	55,933	28,301	27,632	Scotland	31,152	18,805	12,347
Forsyth	233,866	177,983	55,883	Stanly	45,884	41,328	4,556
Franklin	28,641	17,661	10,980	Stokes	31,013	28,988	2,025
Gaston	159,133	139,545	19,588	Surry	57,346	54,602	2,744
Gates	8,244	4,336	3,908	Swain	10,833	7,837	2,996
Graham	6,922	6,546	376	Transylvania	22,444	21,252	1,192
Granville	32,704	19,761	12,943	Tyrrell	3,972	2,479	1,493
Greene	14,774	8,439	6,335	Union	66,788	55,672	11,116
Guilford	311,352	235,022	76,330	Vance	34,671	20,555	14,116
Halifax	55,645	26,919	28,726	Wake	292,326	225,768	66,558
Harnett	56,843	45,995	10,848	Warren	17,278	6,160	11,118
Haywood	44,482	43,508	974	Washington	15,378	9,179	6,199
Henderson	53,231	51,169	2,062	Watauga	30,681	30,372	309
Hertford	24,480	11,717	12,763	Wayne	93,106	60,033	33,073
Hoke	19,376	8,883	10,493	Wilkes	56,889	54,089	2,800
Hyde	5,797	3,873	1,924	Wilson	62,032	41,087	20,945
Iredell	80,301	65,926	14,375	Yadkin	27,904	26,744	1,160
Jackson	26,286	23,493	2,793	Yancey	15,052	14,696	356

Data from N.C. Department of Administration, April 1980, based on 1977 final estimates.

Table 2
CENSUS POPULATION BY RACE: NORTH CAROLINA AND EACH COUNTY, APRIL 1, 1980

<u>Area</u>	<u>Final Total</u>	<u>Preliminary White</u>	<u>Preliminary Nonwhite</u>	<u>Area</u>	<u>Final Total</u>	<u>Preliminary White</u>	<u>Preliminary Nonwhite</u>
North Carolina	5,874,429	4,453,010	1,421,419				
Alamance	99,136	79,619	19,517	Johnston	70,599	56,436	14,163
Alexander	24,999	23,250	1,749	Jones	9,705	5,462	4,243
Alleghany	9,587	9,367	220	Lee	36,718	28,468	8,250
Anson	25,562	13,547	12,015	Lenoir	59,819	36,811	23,008
Ashe	22,325	22,097	228	Lincoln	42,372	38,374	3,998
Avery	14,409	14,226	183	McDowell	35,135	33,415	1,720
Beaufort	40,266	27,430	12,836	Macon	20,178	19,730	448
Bertie	21,024	8,544	12,480	Madison	16,827	16,640	187
Bladen	30,448	18,265	12,183	Martin	25,948	14,334	11,614
Brunswick	35,767	27,273	8,494	Mecklenburg	404,270	291,442	112,828
Buncombe	160,934	145,990	14,944	Mitchell	14,428	14,351	77
Burke	72,504	66,953	5,551	Montgomery	22,469	16,855	5,614
Cabarrus	85,895	73,342	12,553	Moore	50,505	39,393	11,112
Caldwell	67,746	63,739	4,007	Nash	67,153	44,745	22,408
Camden	5,829	3,932	1,897	New Hanover	103,471	80,353	23,118
Carteret	41,092	36,871	4,221	Northampton	22,584	8,824	13,760
Caswell	20,705	11,645	9,060	Onslow	112,784	85,498	27,286
Catawba	105,208	94,974	10,234	Orange	77,055	62,522	14,533
Chatham	33,415	24,316	9,099	Pamlico	10,398	7,100	3,298
Cherokee	18,933	18,275	658	Pasquotank	28,462	17,847	10,615
Chowan	12,558	7,294	5,264	Pender	22,215	13,531	8,684
Clay	6,619	6,569	50	Perquimans	9,486	5,888	3,598
Cleveland	83,435	65,803	17,632	Person	29,164	19,785	9,379
Columbus	51,037	34,406	16,631	Pitt	83,651	54,639	29,012
Craven	71,043	50,408	20,635	Polk	12,984	11,748	1,236
Cumberland	247,160	158,235	88,925	Randolph	91,861	85,610	6,251
Currituck	11,089	9,256	1,833	Richmond	45,481	32,812	12,669
Dare	13,377	12,468	909	Robeson	101,577	39,989	61,588
Davidson	113,162	101,392	11,770	Rockingham	83,426	65,995	17,431
Davie	24,599	21,959	2,640	Rowan	99,186	83,058	16,128
Duplin	40,952	26,835	14,117	Rutherford	53,787	47,102	6,685
Durham	152,785	95,818	56,967	Sampson	49,687	31,861	17,826
Edgecombe	55,988	27,428	28,560	Scotland	32,273	18,746	13,527
Forsyth	243,683	182,647	61,036	Stanly	48,517	42,702	5,815
Franklin	30,055	17,648	12,407	Stokes	33,086	30,574	2,512
Gaston	162,568	141,827	20,741	Surry	59,449	56,321	3,128
Gates	8,875	4,192	4,683	Swain	10,283	7,662	2,621
Graham	7,217	6,826	391	Transylvania	23,417	22,066	1,351
Granville	33,995	18,871	15,124	Tyrrell	3,975	2,418	1,557
Greene	16,117	8,785	7,332	Union	70,380	58,151	12,229
Guilford	317,154	234,579	82,575	Vance	36,748	20,716	16,032
Halifax	55,286	27,559	27,727	Wake	300,833	231,245	69,588
Harnett	59,570	44,861	14,709	Warren	16,232	5,896	10,336
Haywood	46,495	45,550	945	Washington	14,801	8,346	6,455
Henderson	58,580	56,226	2,354	Watauga	31,678	31,110	568
Hertford	23,368	10,285	13,083	Wayne	97,054	64,409	32,645
Hoke	20,383	8,838	11,545	Wilkes	58,657	55,681	2,976
Hyde	,873	3,777	2,096	Wilson	63,132	39,943	23,189
Iredell	82,538	67,942	14,596	Yadkin	28,439	26,969	1,470
Jackson	25,811	22,797	3,014	Yancey	14,934	14,701	233

Prepared from Bureau of Census Publication.

Table 3

A COMPARISON OF 1980 PROJECTION-BASED AND
CENSUS-BASED PROVISIONAL NONWHITE DEATH RATES
FOR 33 COUNTIES HAVING APRIL 1 NONWHITE PROJECTIONS
10% OR MORE BELOW CENSUS COUNTS

County	Death Rates			County	Death Rates		
	Projection-based	Census-based	Percent Decrease		Projection-based	Census-based	Percent Decrease
Ashe	5.0	4.4	12.0	Madison	0.0	0.0	0.0
Avery	10.5	5.5	47.6	Mitchell	0.0	0.0	0.0
Burke	10.5	8.8	16.2	Montgomery	12.9	11.2	13.2
Cherokee	12.1	10.6	12.4	Moore	13.0	11.6	10.8
Cleveland	10.4	9.2	11.5	Onslow	3.6	3.1	13.9
Cumberland	5.8	4.9	15.5	Person	10.2	8.4	17.6
Dare	8.9	7.7	13.5	Randolph	11.8	10.4	11.9
Davidson	8.5	6.9	18.8	Rockingham	11.3	9.8	13.3
Davie	7.3	5.7	21.9	Rutherford	10.7	8.2	23.4
Franklin	10.2	9.0	11.8	Stanly	12.1	9.5	21.5
Gates	14.1	11.7	17.0	Stokes	7.4	6.0	18.9
Granville	11.2	9.6	14.3	Surry	8.0	7.0	12.5
Greene	8.2	7.1	13.4	Transylvania	8.4	7.4	11.9
Harnett	12.2	9.0	26.2	Vance	12.4	10.9	12.1
Henderson	16.5	14.4	12.7	Watauga	6.5	3.5	46.2
Jones	8.0	7.1	11.3	Yadkin	17.2	13.6	20.9
Macon	14.0	8.9	36.4				

population-based rates using the middle-year census counts as applicable to each of the three years. The volume will resume 5-year rates whenever corrected pre-1980 population bases are available or will build up to 5-year rates in data year 1983, whichever comes first.

- The 1981 Volume 2 will resume standard format with 3-year rates (1979-81) replacing 5-year rates. If corrected 1979 population bases are unavailable, the volume will use middle-year 1980 census counts as above.
- The 1982 and 1983-and-forward Volume 2's will show 4-year and 5-year rates respectively.

Final Comments

For the social scientist, population-based rates are a cornerstone to trend analysis. As one examines trends in health, education, corrections, etc., it is natural to question the completeness and validity of these data. It is also natural to direct these questions at the social event being measured, i.e., the numerator of the rate. Consideration is seldom given to the quality of the measurement of the population at risk—the denominator. Denominators based on accurate census data are crucial, especially if one is examining small area statistics, category-specific rates or adjusted rates. It is not sufficient to judge the quality of social statistics by the quality of the data system that generates the numerator. It is not sufficient for agencies generating social statistics to report on the quality of their data only. It is necessary for both researchers and statistics agencies to examine the effects of errors in census counts and projections of those counts on any trend analysis being undertaken. During the next year, SCHS will endeavor to make the reader aware of real changes vs. artificial changes in the health status of North Carolinians.

ERRATUM

Statistical Primer Vol. 1 No. 1 p. 7.
Comparison 2 under Indirect. Change
standard to study.



NORTH CAROLINA
Department of Human Resources
Division of Health Services
State Center for Health Statistics
P.O. Box 2091, Raleigh, N.C. 27602

Francine J. Ewing
State Demographer
State Budget & Management
Administration Bldg.